

Modelling

Regression task since all the predictors and the response are continuous variables.

A **gradient boosted decision tree** was used to solve the regression task.

A more simple and interpretable model (**linear regression**) was used as a baseline to compare the performance of the gradient boosting model.

Lasso and ridge methods were applied, multicollinearity wasn't a problem and the improvement in terms of test error was minimal so I decided in favor of the more simple linear regression.

Software



Data manipulation and visualization

Caret package.

Uniform interface to standardize common modelling tasks. In particular for cross-validation

XGBoost

Optimized distributed library for a gradient boosting framework

xgboostExplainer package.

Tools to interpret the results of XGBoost

